Multivariate Time Series Analysis and Forecasting on Air Quality Index with Industrial Statistics: A Case Study of Bangladesh

MD Shafikul Islam\textsuperscript{1*}, Azmine Tousik Wasi\textsuperscript{2*}

\textsuperscript{1}Department of Industrial and Production Engineering, SUST, Sylhet, 3114
\textsuperscript{2}Department of Industrial and Production Engineering, SUST, Sylhet, 3114

Email:
\textsuperscript{1}shafikul37@student.sust.edu \textsuperscript{2}azmine32@student.sust.edu

*Corresponding author [MD Shafikul Islam, Azmine Tousik Wasi]

Keywords:
- Air Quality Index
- Multivariate Analysis
- Industrial Intensity Index
- Motor Vehicle Registered
- Facebook Prophet

Abstract: The Air Quality Index (AQI) analyzes all the observed particles and pollutants; converts them to a single scale number to represent how fresh or polluted the air is in a particular location. AQI values correlated with many factors including environment, manufacturing, industrialization, transportation, and weather. In this research, a multivariate time series analysis of the Air Quality Index was conducted with other statistics such as the Industrial Intensity Index, Manufacturing Value Added per capita, Motor Vehicles, and CO\textsubscript{2} Emission. These variables are analyzed concerning time to identify their relationship with AQI by the Pearson correlation method. According to our findings, Manufacturing Value Added and Total Motor Vehicles Registered in a Year are highly correlated with AQI, with correlation values of 0.84 and 0.85, respectively; defining AQI is primarily affected by pollution from transportation and manufacturing. The Prophet time series model was also used to learn trends and forecast future AQI values. According to the model, the AQI value rises from 2016 to 2022 but falls sharply in 2020; the average AQI value is predicted to be 178.89 in 2023, which is higher than any previous year's value.

1. Introduction
The Air Quality Index (AQI) measures the quality of the ambient air concerning the applicable international air quality standard. This is a thorough index that considers multiple pollutants at once and can reflect the overall status of the ambient air quality of the experiment area. (Baldwin R, et al. (2007). Furthermore, the Air Quality Index (AQI) concentrates on the health effects of air pollution that may appear hours or days after inhaling polluted air.
Air quality was identified as the top environmental concern for human health worldwide in the Environmental Health Policy Objectives of the 2022 Environmental Performance Index (EPI) (Wendling, et al. (2022)). According to the World Air Quality Report 2020, Bangladesh has the worst air quality in the world, with Dhaka ranking second among the most polluted cities (Ahmed H.U. (2020)). Not only Bangladesh, but the entire world, is suffering from the effects of air pollution. Approximately 93% of children also breathe toxic air (Anik S.S.B (2020)).

The UNIDO competitive Industrial Performance (CIP) Index, developed primarily for developing countries, benchmarks the country's competitive industrial activity against the backdrop of liberalization and globalization. Manufacturing value added and the Industrial intensity index are two important components of it. MVA per capita is a proxy for a country's level of industrialization and is deflated by population to account for the size of the economy. The intensity of industrialization is calculated by taking the simple average of the share of MVA in GDP and the share of medium and high-technology (MHT) activities in MVA (Komninos N. (2005)). Besides, CO2 emissions are the primary cause of global climate change. It is widely acknowledged that to avoid the worst effects of climate change, the world must urgently reduce emissions. (Ritchie H, et al. (2020)).

Several studies have been conducted to demonstrate the significance of the Air Quality Index. According to Islam and Salam, gaseous pollutants contribute significantly to Dhaka's particulate pollution (2000 and 2003). Ahmed and Hossain investigated particulate pollution in Dhaka and Chittagong, Bangladesh's two largest cities (2008). With time series modeling, Islam and Sharmin predicted the air quality of the Sylhet and Dhaka divisions (2020). Azad, Rahman, and collaborators conducted a study in Bangladesh to examine the changes in air quality and COVID-19 transmission in various lockdown scenarios (2021). However, the relationship between the Air Quality Index and other industrial and environmental parameters is not well understood. Most importantly, making decisions requires forecasting the air quality using other parameters because the impact of COVID-19 has almost been neutralized.

Given the significance of the Air Quality Index, the goal of this study is to identify the overall pattern of the Air Quality Index concerning other industrialization statistics in Bangladesh, such as the Industrial Intensity Index, Manufacturing Value Added, and Motor Vehicle Registered. We have also considered one more Environmental parameter, CO2 emission to find the co-relation with the Air Quality Index.

2. Methodology

Basic Overview of the method

The purpose of our research is to determine the relationship between the Air Quality Index and other factors such as industrial, environmental, and transportation statistics in Bangladesh. To accomplish this goal, data was gathered from various sources.
AQI values are analyzed to detect underlying monthly, and yearly trends. Data is smoothened into yearly moving average to determine those trends, and to study the behavior of the data. For developing a prediction model, Facebook prophet time series modeling was used. The model is highly efficient and able to measure different trends and forecast future AQI values based on the previous data.

For co-relation analysis, all the factors including Bangladesh Industrial Performance Index (III), Manufacturing Value Added per Capita (MVA), Total Motor Vehicle Registered (MVR) and \( CO_2 \) Emission rate (CO2), and Air Quality Index (AQI) values are scaled to mean 0 and standard deviation 1, for better processing in the next steps. The p-value threshold for statistical significance of correlation was fixed at \( p \leq 0.05 \). Multivariate analysis was conducted on the scaled data to analyze their relationship and dependence. Later, Pearson method was used to determine correlation coefficients. Finally, some statistical tests were performed to verify if the co-relations are statistically significant and responsible.

**Figure 1: Methodology Overview: An Overview Diagram Illustrating the Flow of Analysis**

Based on the significance test results and the output of the Facebook prophet time series model, current and possible future condition of Bangladesh in terms of Air Quality Index and Industrial Statistics was evaluated and necessary next steps were suggested.

**Data Description**

AQI data are obtained from (https://waqi.info/) on October 1, 2022. Data on Bangladesh's Manufacturing Value Added and CO2 emissions are accessed from (https://datacatalog.worldbank.org). To retrieve data for the Industrial Intensity Index and Motor Vehicle Registration, we created a web scraping algorithm using the Python package BeautifulSoup (Richardson, L. (2007)).
Pearson Correlation

The Pearson correlation method is the most used approach for numerical variables. It assigns a value between -1 and 1 for presenting the correlation with two variables where 1 represents the maximum positive correlation and -1 represents the maximum negative correlation. A positive correlation indicates that if variable X increases, variable Y will also increase, while a negative correlation indicates that if X increases, Y decreases (Boslaugh & Watters, 2008). The equation is given below:

\[
 r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}
\]

Facebook Prophet Model

Facebook Prophet is a method for forecasting time series data based on an additive model with yearly, weekly, and daily seasonality, as well as holiday effects. Prophet is resistant to missing data and trend shifts, and it typically manages outliers well (Letham, 2017). At the core of this modeling, the output function \( Y(t) \) is the addition of growth \( g(t) \), seasonality \( s(t) \), holiday \( h(t) \), and error \( e(t) \).

\[
 Y(t) = g(t) + s(t) + h(t) + e(t)
\]

3. Result and Data Visualization

AQI: Time Series Analysis of AQI and Forecasting with Facebook Prophet

The Air Quality Index (AQI) is a measure of concentrations of different air pollutants. It assigns numbers on a scale of 0 to 500 showing current air quality. It also can be used to predict when air quality will be unhealthy by time series analysis with enough data. From the time series analysis of AQI, it can be noticed that there was always a steady rise in AQI value at the starting and ending part of every year.

Figure 2: AQI – Original Series and One-Year Moving Average
As per figure 2, the AQI values (Original Series) follow a trend of higher AQI values at the start and end of the year and stable in the middle. And, in the One Year Moving Average (Moving Average (Yearly)) graph, we can notice a subtle decrease in AQI in 2020, pointing to the COVID-19 timeline.

Figure 3 demonstrates that in 2023, it is expected to have some high outliers at the beginning of 2023 with an average AQI of 178.89. Also, the average AQI value is rising regularly starting in 2020. It is predicted to rise by 2.22% and 9.07% in 2022 and 2023 respectively from previous years, raising a huge concern for Bangladesh.

**Multivariate Time Series Analysis**

In Figure 4, a visualization was prepared for the 4 factors along with AQI in a line chart. In visual inspection, Manufacturing Value Added per Capita (MVA) and Total Motor Vehicle Registered (MVR) demonstrated close and positive co-relation. Industrial Intensity Index (III) also followed a similar trend except in 2019.
Pearson Correlation and Statistical Significance

The Pearson method was used to determine the correlation of all the factors. The results are shown in Table 1 and visualized as a heatmap in Figure 5, where darker regions denote higher correlation.

<table>
<thead>
<tr>
<th></th>
<th>AQI</th>
<th>CO2</th>
<th>MVA</th>
<th>III</th>
<th>MVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQI</td>
<td>1.00</td>
<td>0.387</td>
<td>0.847</td>
<td>0.272</td>
<td>0.856</td>
</tr>
<tr>
<td>CO2</td>
<td>0.387</td>
<td>1.000</td>
<td>0.571</td>
<td>-0.249</td>
<td>-0.142</td>
</tr>
<tr>
<td>MVA</td>
<td>0.847</td>
<td>0.571</td>
<td>1.000</td>
<td>0.221</td>
<td>0.617</td>
</tr>
<tr>
<td>III</td>
<td>0.272</td>
<td>-0.249</td>
<td>0.221</td>
<td>1.000</td>
<td>0.428</td>
</tr>
<tr>
<td>MVR</td>
<td>0.856</td>
<td>-0.142</td>
<td>0.617</td>
<td>0.428</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 1: Correlation Matrix of Factors and AQI

The figure shows that Manufacturing Value Added per Capita (MVA) and Total Motor Vehicle Registered (MVR) illustrate a very higher correlation with AQI; 84.7% and 85.6% respectively verifying the visual inspection. On the other hand, CO2 Emission and Industrial Intensity Index (III) show weak but positive correlation; 38.7% and 27.2% respectively. In a two-tailed Pearson significance test, both the correlations of Manufacturing Value Added per Capita (MVA) and Total Motor Vehicle Registered (MVR) to AQI are recognized as statistically significant maintaining p ≤ 0.05.

4. Discussion

This study correlated the air quality index with other statistics to determine the actual concerning area in order to reduce the increasing trend of the air quality index. The Pearson correlation test revealed that the relationship between Manufacturing Value Added and Motor Vehicle Registered is statistically significant. This research will help to develop an integrated approach to recovering from the negative effects of the environment and will lead the way for relevant recommendations to improve the country's air quality. However, the correlation between the Industrial Intensity Index and the Air Quality Index is not statistically significant. Despite rapid industrialization, Bangladesh has a good reputation for establishing its Green Industry (Star Business Report, 2019).

<table>
<thead>
<tr>
<th>Statistics Type</th>
<th>Correlation value</th>
<th>P value (Two-tailed test)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial Intensity Index</td>
<td>0.27</td>
<td>0.601</td>
<td>Not significant</td>
</tr>
<tr>
<td>Manufacturing Value Added</td>
<td>0.84</td>
<td>0.033</td>
<td>Significant</td>
</tr>
<tr>
<td>Motor Vehicle Registered</td>
<td>0.85</td>
<td>0.029</td>
<td>Significant</td>
</tr>
<tr>
<td>CO2 Emission</td>
<td>0.38</td>
<td>0.44</td>
<td>Not significant</td>
</tr>
</tbody>
</table>

Table 2: Correlation, P value, and statistical significance of different statistics and AQI index
Air quality forecasting was done using the Facebook Prophet model, which is a useful technique as it takes seasonality, holidays, and other errors into account. The ability to predict future air quality with 95% confidence intervals shows how accurately the models measure their parameters. However, for more precise forecasting, some unpredicted variables (rainfall, temperature, etc.) could be incorporated into the model. According to the study’s findings, air quality in the country remains unhealthy or extremely unhealthy during the dry season (December-March). The severity of most construction and urbanization grows, and the level of concern during this time grows more serious.

5. Conclusion

The purpose of this research is to determine the overall pattern of the Air Quality Index concerning other industrialization statistics in Bangladesh and find a credible forecasted AQI value. This study finds that rather than the Industrial Intensity Index, Total Motor Vehicle registered, and Total Manufacturing Value added is the actual concerning area for increasing trend in the air quality index. It will be beneficial in providing objective guidelines and reliable decision-making regarding the condition of the air quality index in the future. As our model predicts that contamination will increase in the future, identifying and mitigating the contributions of each factor in increasing the air quality index will be an excellent way to address the critical issue.
References


Boslaugh, S., & Watters, P. A. 2008. Statistics in a Nutshell: A Desktop Quick Reference (In a Nutshell (O’Reilly)).


